

Praktische Ziehung von Zufallsstichproben für Telefon-Surveys

Schnell, Rainer

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Schnell, R. (1997). Praktische Ziehung von Zufallsstichproben für Telefon-Surveys. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 40, 45-59. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-121771>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Praktische Ziehung von Zufallsstichproben für Telefon-Surveys

von Rainer Schnell ¹

Zusammenfassung

Es werden verschiedene Ansätze für Stichprobenpläne vorgestellt, mit denen aus Telefon-CD-ROMs Zufallsauswahlen von Telefonnummern durchgeführt werden können. Die dabei auftretenden praktischen Probleme und deren Lösungen werden diskutiert. Hierzu werden Hilfsprogramme in der String-Verarbeitungssprache AWK verwendet.

Abstract

Sample designs for telephone surveys using German CD-ROMs of telephone numbers are discussed. The practical problems during actual selection are solved by small programs using the string processing language AWK.

"Well, last week we showed you how to become a gynaecologist. And this week on 'How to do it' we're going to show you how to play the flute (...) How to play the flute. (picking up a flute) Well here we are. You blow there and you move your fingers up and down there." (*Chapman* 1990:63-64)

"- Auswahlbasis auf der ersten Stufe sind sogenannte Knoten. Darunter versteht man Ortsvermittlungsbereiche bzw. Ortsnetze, die über keine Ortsvermittlungsbereiche verfügen. Davon gibt es etwa 6.000 in ganz Deutschland. Daraus wird eine Zufallsauswahl gezogen.

- In einem weiteren Schritt wird dann für jeden ausgewählten Knoten eine Nummer ausgewählt." (*Fuchs* 1994:162)

Da in der Bundesrepublik kein akademisch basiertes Erhebungsinstitut existiert, werden alle bundesweiten Surveys für face-to-face-Interviews durch die kommerziellen Datenerhebungsinstitute durchgeführt². Telefoninterviews bieten hingegen Universitäten prinzipiell die Möglichkeit, lokale und auch bundesweite Erhebungen ohne Inanspruchnahme kom

¹ Anschrift: Dr. **Rainer Schnell** ist Professor an der Fakultät für Verwaltungswissenschaft, Universität Konstanz, Postfach 5560, 78434 Konstanz, email: Rainer.Schnell@Uni-Konstanz.De

² Zu den damit verbundenen Problemen vgl. u.a. *Schnell* (1997:253-254).

merzieller Institute durchzuführen. Der praktischen Umsetzung stehen vor allem zwei Probleme entgegen:

1. Es existiert anscheinend keine verfügbare CATI-Software in der Public Domain oder zu Preisen, die im akademischen Bereich noch finanziell tragbar wären³.
2. Es wurden bislang kaum Stichprobenpläne für Telefonsurveys in der Bundesrepublik publiziert.

Das erste Problem läßt sich z.B. durch den Einsatz jeweils für den speziellen Zweck angefertigter Einzelprogramme umgehen⁴. Häufiger wird bei akademischen Surveys auf die CATI-Technologie verzichtet und statt dessen traditionell mit Interviewern gearbeitet, die am Telefon einen gedruckten Fragebogen ausfüllen⁵.

Das zweite Problem ist zumindest für bundesweite Stichproben schwieriger zu lösen. Ideen für Stichprobenpläne und Ansätze prinzipiell praktikabler Ziehungsverfahren wurden in der Bundesrepublik bisher nur für Auswahlen anhand der gedruckten Telefonbücher publiziert⁶. Es ist leicht einzusehen, daß die damit verbundenen praktischen Probleme aufgrund des hohen Arbeitsaufwandes beachtlich sind. Nach dem Erscheinen von CD-ROMs, die die Telefonnummern der BRD enthalten, wird die Durchführung eines Auswahlverfahrens für Telefonsurveys nunmehr häufig fälschlich als trivial angesehen. Praktische Probleme der Durchführung von Stichprobenziehungen zeigen sich aber erst bei ihrer tatsächlichen Realisierung. Einige der auftretenden Schwierigkeiten sollen hier dargestellt und Möglichkeiten zur ihrer Überwindung beschrieben werden.

Telefon-CDs als Auswahlgrundlage

Für die Bundesrepublik werden CD-ROMs mit den in den Telefonbüchern eingetragenen Nummern der Teilnehmer derzeit von drei verschiedenen Herstellern angeboten⁷. Um als Auswahlgrundlage brauchbar zu sein, muß die Datenbank die Möglichkeit bieten, alle Datensätze exportieren zu können⁸. Da sowohl die Telekom-CD als auch die Tele-Info-CD

3 Kommerzielle Systeme wie BLAISE, CI-3, Surveycraft oder in2itive liegen in der Anschaffung derzeit effektiv bei ca. 1000 bis 2000 DM pro Interviewerstation.

4 Leider fehlen diesen Programmen dann eine große Zahl derjenigen Features, die CATI-Interviews eigentlich interessant machen, wie z.B. das Call-Management.

5 Die wohl beste und detaillierteste Durchführungsanweisung für solche Surveys findet sich bei *Lavrakas* (1987).

6 Prinzipielle Möglichkeiten beschreibt *Zeh* (1987), technische Details in größerem Umfang finden sich bisher lediglich bei *Frey, Kunz und Lüschen* (1990:83-98).

7 DeTeMedien (1996): Telefonbuch für Deutschland (CD-ROM), Frankfurt.
Tele-Info Verlag (1996): Xi-Deutschland (CD-ROM), Garbsen. TopWare (1996): D-Info 3.0 (CD-ROM), Dorsten.

8 Dies ließe sich nur umgehen, wenn die Hersteller bereits in der Abfragesoftware der CDs die Möglichkeit einer Zufallsauswahl vorsehen würden.

ein Exportlimit von maximal einigen Hundert Nummern besitzt, kommt für die praktische Ziehung von Telefonstichproben nur die D-Info 3.0 CD in Frage⁹.

Praktische Probleme der Dateigröße

Bei der Verwendung der D-Info 3.0 CD als Auswahlgrundlage entstehen eine Reihe technischer Probleme durch die Größe der resultierenden Dateien. Da die CD mit einem nicht dokumentierten Kompressionsalgorithmus beschrieben wurde, sind die Daten nicht direkt lesbar. Man kann Datenbestände allerdings exportieren. Die Software erlaubt als Fileformat Dbase- oder ASCII-Files. Dbase-Format besitzt den Vorteil, daß es direkt von anderen Programmen wie z.B. SPSS gelesen werden kann. Das Export-Format sieht aber eine fixe Recordgröße von 281 Byte vor. Bei 34.3 Millionen Einträgen ergäbe sich daher eine Dateigröße von 9.2 Gigabyte. DOS kann aber nur Platten bis ca. 2 Gigabyte verwalten. Das Dbase-Format eignet sich folglich nur für Files bis ca. 7.5 Millionen Einträgen. Werden mehr Einträge benötigt, muß auf das ASCII-Format ausgewichen werden. Verwendet man das ASCII-Format unter DOS, so sieht D-Info ein "embedded"-Stringformat vor: Die einzelnen Felder (Name, Straße, Ort, Nummer) sind jeweils in Anführungsstrichen eingeschlossen und durch Semikolon getrennt. Dieses Format führt für den gesamten Datenbestand zu einer Dateigröße über 2 Gigabyte. Unter DOS kann der gesamte Datenbestand daher nicht exportiert werden. Unter Windows erlaubt D-Info hingegen die Veränderung des ASCII-Exportformats. Verzichtet man auf die Einbettung in Anführungsstriche, so reduziert sich der Speicherbedarf um ca. 262 Megabyte auf 1.96 Gigabyte. Es empfiehlt sich hierbei einen Feldtrenner zu verwenden, der im normalen Datenbestand nicht vorkommt, so z.B. den senkrechten Strich (|). Diese Datei enthält ca. 34.324.400 Records¹⁰. Dateien dieser Größe lassen sich mit den meisten Standardprogrammen nicht mehr verarbeiten¹¹.

Hilfsprogramme für den Umgang mit dem Datenbestand

Das Exportprogramm der CD exportiert vier Felder (Name, Straße, Ort, Nummer); für die meisten Auswahlverfahren bei Telefonsurveys werden hiervon nur Nummer und Name benötigt. Da schon allein die kleinere Dateigröße den Umgang mit diesen Dateien wesentlich erleichtert, sollte man durch geeignete Filterprogramme die nicht benötigten Felder löschen.

9 Die Exportlimits lassen sich zwar prinzipiell durch verschiedene Möglichkeiten (z.B. Makros, Dekodierungsprogramme, Disassemblieren und Patchen der Retrievalsoftware) umgehen, allerdings wäre der Arbeitsaufwand trotzdem vergleichsweise hoch. Weiterhin würden solchen Verfahren klar gegen die Nutzungsbedingungen der CDs verstoßen.

10 Die genaue Zahl hängt davon ab, welches Kriterium man verwendet um fehlerhafte Einträge bereits im ersten Lesevorgang zu eliminieren.

11 Bemerkenswerterweise gilt dies sogar für eine Reihe von Compilern. So lassen sich z.B. Files über 500 Mbyte anscheinend in Turbo-Pascal 7.0 (Borland 1992) nicht ansprechen. Die kompilierten PASCAL-Programme brechen mit irreführenden Fehlermeldungen ("file not found") ab.

Filterprogramme werden für eine Reihe weiterer Aufgaben bei der Durchführung von Auswahlverfahren benötigt.

Programmiert man einige Datenprüfungen mit, so lassen sich solche Filterprogramme in einer allgemeinen Programmiersprache wie FORTRAN, C++ oder PASCAL mit weniger als 100 Programmzeilen realisieren. Die Programmentwicklung ist allerdings wesentlich einfacher, wenn man eine auf Stringverarbeitung spezialisierte Programmiersprache wie Perl (*Wall, Schwartz* 1993) oder AWK (*Aho, Kernighan und Weinberger* 1988) verwendet¹². Mit diesen Sprachen lassen sich solche Aufgaben mit wenigen Zeilen Programmcode erfüllen. Perl eignet sich auch für größere Programme, ist aber etwas schwerer zu erlernen als AWK. AWK ist eine ideale Sprache für kleine Gelegenheitsprogramme, die nur wenige Male benötigt werden und deren Effizienz keine Rolle spielt¹³. Die im folgenden berichteten Prozeduren wurden entsprechend zum größten Teil mit AWK-Programmen durchgeführt, die selten mehr als 10 Zeilen beanspruchen¹⁴. Beispiele für AWK-Programme bzw. "Oneliner" finden sich unten¹⁵. Falls in ASCII-Files nur einzelne Zeilen gelöscht oder gesucht werden sollen, empfiehlt sich neben AWK die Verwendung des Unix-Programms "grep", das auch in verschiedenen Varianten für DOS zur Verfügung steht. Weiterhin benötigt man für viele Aufgaben, wie z.B. der Bestimmung der Zahl der Doppeleinträge, ein leistungsfähiges Sortierprogramm, das unter DOS leider nicht zum Betriebssystem gehört¹⁶.

Schließlich tritt ein besonderes Problem dadurch auf, daß nur wenige DOS und Windows-Editoren Files mit einer Größe von mehreren hundert Megabyte bearbeiten können¹⁷. Die für die Auswahl notwendigen Aufgaben lassen sich für den gesamten Datenbestand entwe-

12 Compiler bzw. Interpreter für beide Sprachen sind im Internet kostenlos erhältlich. AWK gehört bei UNIX-Systemen zu Lieferumfang. Eine DOS-Version von AWK (Autor: **Rob Duff**) findet sich im SIMTEL-Archiv. SIMTEL ist z.B. über <ftp://ftp.uni-heidelberg.de> verfügbar. AWK liegt als AWK320.ZIP im Verzeichnis "simtelnet/msdos/txtutil". Weiterhin gibt es ein GNU AWK (GAWK), das auch für DOS kompiliert wurde. Diese schnelle Version findet sich u.a. unter "<http://www.leo.org/pub/comp/platforms/pc/gnuish/>".

13 Den Hinweis auf die Möglichkeiten von AWK beim Datenmanagement verdanke ich Professor **Wolfgang Sodeur**.

14 Da sich AWK aber bei Programmfehlern fast immer auf die Mitteilung "Syntax Error" ohne weitere Hinweise beschränkt, kann die Entwicklung der ersten eigenen Programme etwas länger beanspruchen als man ursprünglich glaubt.

15 Bei den Beispielprogrammen sind Zeilen, die mit "#" beginnen, Kommentarzeilen. In der ersten Kommentarzeile jedes Programms findet sich hier der Programmname. Die Beispielprogramme müssen in jeweils einem ASCII-File mit diesem Namen gespeichert werden, so z.B. "DUPS.AWK". Um das Programm auszuführen muß AWK installiert sein. Unter DOS erfolgt die Installation durch Kopieren eines Files (AWK.EXE) in ein beliebiges Verzeichnis im Suchpfad. Der Aufruf erfolgt dann mit "AWK DUPS inputfile > outputfile".

16 Im SIMTEL Verzeichnis "txtutil" (siehe oben) finden sich eine Reihe nützlicher DOS Hilfsprogramme, so z.B. auch das Sortierprogramm MSORT (MSORT115.ZIP) von **Martin Katz**. MSORT sortiert auch Dateien, deren Größe über hundert Megabyte liegt, wenn auch - natürlich - mit entsprechender Dauer.

17 Ein solcher Editor ist z.B. MEL von American Cybernetics Inc. Eine Version findet sich im SIMTEL-Directory "simtelnet/msdos/editors" als "MELITE.ZIP".

der problemlos mit einer leistungsfähigen Unix-Maschine oder mit größeren Problemen mit einer Unix-Workstation bzw. einem PC mit den erwähnten Hilfsprogrammen durchführen. Für die Bearbeitung von Teilmengen des Datenbestandes, wie z.B. einzelner Gemeinden, reicht hingegen schon ein kleinerer PC mit Standardsoftware wie z.B. SPSS oder auch EXCEL aus.

Übersicht 1: Sechs verschiedene vollständige AWK-Programme zur Listenverarbeitung.

```
# DUPS.AWK: count the number of duplicate input lines
# output: only dups, first field: number of replications
BEGIN { old="_Z@1_Z@1Z_Z@1";c=1 }
{
  if ($0==old) { c=c+1 }
  else { if (c > 1) {print c,old; c=1}; old=$0 }
}
END { if (c > 1) {print c,old } }

-----

# UNIQUE.AWK: report only unique input lines, omit duplicates
BEGIN { old="_Z@1_Z@1Z_Z@1" }
{ if ($0 != old) { print $0; old=$0 } }

-----

# LINES.AWK: number of records
END {printf("%s%d\n","Number of Records: ",NR) }

-----

# FREQF1.AWK: frequencies of the first field
{
  a=$1
  count[a]++
}
END { for (i in count)
      print i ": ", count[i]
}

-----

# ONLYF1.AWK: read comma delimited, embedded strings.
# Kill ", keep only field 1
BEGIN {FS=","}
{ gsub("\\"", "", $0); print $1 }

-----

# PREFREQ.AWK: count dial prefix in a sorted file
BEGIN {FS="-";old="0000";c=1}
{
  if ($1==old) { c=c+1 }
  else { if (c > 1) {print old,c; c=1}; old=$1 }
}
END { print old,c }
```

Probleme der Datenbasis

Trivialerweise enthalten die Telefon-CDs ebenso wie die Telefonbücher lediglich die Einträge derjenigen Anschlüsse, deren Besitzer den Eintrag ins Telefonbuch wünschen und

deren Anschluß nicht kürzlich erfolgte. Da beide Mechanismen nicht zu zufälligen Ausfällen führen¹⁸, sollte **keine** Auswahl aus der Liste (dem Telefonbuch oder der CD) erfolgen, sondern ein Eintrag in der Liste als Ausgangszahl für das "Randomized Last Digit"-Verfahren (RLD) verwendet werden, bei dem zur Ausgangszahl eine gleich verteilte Zufallszahl zwischen 0 und 9 addiert wird (ein Beispiel für ein entsprechendes AWK-Programm findet sich unten). In der Praxis wird hierauf allerdings häufig verzichtet: Zum einen aus Kostengründen (da viele der generierten Nummern keine Privatanschlüsse darstellen), zum anderen um sich Diskussionen mit aufgebrachten Zielpersonen, deren Nummer absichtlich nicht eingetragen wurde, zu ersparen.

Übersicht 2: AWK-Programm RLD.AWK: Addiert eine Zufallszahl zu einer Liste mit Telefonnummern

```
# RLD.AWK: add a random digit to a list of phone numbers
BEGIN {srand()}
{
  i=index("-")
  if (i==0)
    { print "no prefix in following input line: ", $0; exit }
  pre=substr($0,1,i)
  tel=substr($0,i+1)
  gsub("-", "", tel)
  r=int(rand()*10)
  printf("%s%d\n",pre,tel+r)
}
```

Die exportierten Dateien enthalten Sondernummern, deren Vorwahl erkennen läßt, daß sie kaum als Privatanschluß in Frage kommen, so z.B. 0130, 0180, 0190. Es empfiehlt sich, diese Nummern durch ein Filterprogramm zu löschen. Es stellt sich die Frage, ob dies auch für die Mobil-Telefonnummern (0161, 0171, 0172, 0177) gilt. Übliche Praxis in der BRD durch die kommerziellen Institute ist es, alle diese Nummern aus der Auswahl auszuschließen. Die Begründung hierfür liegt darin, daß vermutlich fast alle Besitzer von Mobil-Telefonen auch über einen stationären Anschluß erreicht werden können und sie daher eine höhere Auswahlwahrscheinlichkeit besäßen, wenn man sie nicht ausschloesse.

Dies ist vermutlich für die neuen Bundesländer zumindest zur Zeit nicht in gleichem Umfang gegeben wie für die alten Bundesländer. Eine Auszählung der entsprechenden Vorwahlen nach Bundesländern in der DInfo 3.0 CD zeigt so deutlich höhere Anteile von Sonder- vorwahlen (Mobilanschlüsse + 0130, 0180, 0190) in den neuen Bundesländern als in den alten Bundesländern (zwischen 2.8% und 3.8% in den neuen, zwischen 0.9% und 2.7% bzw. 3.7% mit dem Ausreißer Rheinland-Pfalz in den alten Bundesländern). Dieses Ungleichgewicht betrifft aber das generelle Problem der Anschlußdichte in den neuen Bundesländern

18 Eine Liste der amerikanischen Literatur zu diesem Problem findet sich bei *Hüder* (1996).

und soll daher hier nicht weiter verfolgt werden. Generell ließe sich das Problem der Mobilanschlüsse nur durch die tatsächliche Berücksichtigung bei der Auswahl und der expliziten Frage nach weiteren stationären Anschlüssen klären. Abgesehen von der vermutlich höheren subjektiven Belastung der potentiellen Zielperson bei Mobilanschlüssen durch den Anruf und den daraus resultierenden Verweigerungen sowie des vermutlich doch geringen Anteils der nicht über stationäre Anschlüsse erreichbaren Zielpersonen sprechen gegen die Berücksichtigung der Mobilanschlüsse die weit höheren Gesprächsgebühren. Daher sollte das Filterprogramm auch die Mobiltelefonnummern vor der Auswahl löschen.

Die vom D-Info 3.0 Exportprogramm ausgegebenen Felder für Telefonnummern enthalten nicht nur Telefonnummern, sondern auch eine Reihe von Sondersymbolen wie z.B. "ISDN", "Q", "NEU" und "+". Diese Symbole sollten vom Filterprogramm gelöscht werden. Prüft man die verbleibenden Zeichenketten daraufhin, ob sie Telefonnummern darstellen, so entdeckt man eine Reihe von Fehlern. Hierzu gehören mehrere Telefonnummern in einem Feld, die durch Komma getrennt sein können oder auch nicht. In anderen Fällen stellen die Zeichen in diesem Feld keine Nummern dar, sondern Strings wie z.B. "Beratung". Insgesamt finden sich ca. 1500 solcher Fehler im gesamten Datenbestand. Dies ist zwar ein verschwindend geringer Anteil (ca. 0.0045%!), verhindert aber die problemlose Übernahme in automatische Wählprogramme¹⁹.

Um die Erhebungskosten zu senken, wäre eine eindeutige Klassifikation der Telefonnummern nach Privat- und Geschäftsanschluß nützlich. Eine solche liegt im Datenbestand nicht vor.

Man kann natürlich die Strings im Feld mit dem Namen des Anschlußinhabers klassifizieren²⁰. Hierzu benötigt man eine Liste von Strings, die auf einen Geschäftsanschluß hindeuten. Einen vorläufigen Vorschlag für eine solche Liste zeigt die Abbildung 1. Legt man diese Liste zugrunde, so lassen sich je nach Bundesland zwischen 7.5% (NRW) und 12.2% (Sachsen) als Geschäftsanschluß klassifizieren²¹. Verwendet man keine "randomized last digit"-Stichproben, so können die entsprechenden Zeilen vom Filterprogramm gelöscht werden.

19 Ein Filterprogramm sollte daher prüfen, ob der Telefonnummernstring tatsächlich nur Ziffern und "-" enthält, wobei keine zwei "-" aufeinander folgen dürfen.

20 Hierbei muß beim Filterprogramm beachtet werden, daß Folgenummern bei Geschäftsanschlüssen häufig durch leere Namensfelder oder durch ein spezielles Symbol (z.B. "o") nach einem gültigen Namen gekennzeichnet sind.

21 Diese naheliegende Idee findet sich auch bei *Marhenke* (1996). Allerdings gibt *Marhenke* nicht die Liste an; er klassifiziert anhand seiner Kriterien für Kassel (und für die D-Info 2.0 CD) 1.6% der Anschlüsse als Geschäftsnummern (eigene Berechnung, R.S.); an anderer Stelle gibt er "mindestens 7%" an. Anhand der Kriterien der hier vorgeschlagenen Liste werden 9.2% der Anschlüsse als Geschäftsnummer klassifiziert.

Abbildung 1: Die verwendeten 75 Schlüsselworte zur Identifikation vermutlicher Geschäftsanschlüsse

'AG', 'e.V.', 'Fax', '&', 'Automobil', 'BTX', 'Btx', 'büro', 'Club', 'club', 'Center', 'center', 'Dienst', 'dienst', 'emeinschaft', 'ermietung', 'GdBR', 'Gesellschaft', 'Immobilien', 'Kfz', 'mbH', 'MBH', 'Mobiltelefon', 'Service', 'service', 'Studio', 'team', 'technik', 'Telefax', 'Verband', 'verband', 'verein', 'Verein', 'Versicherung', 'versicherung', 'Vertrieb', 'vertrieb', 'KG', 's. unter', 'o', 'praxis', 'Praxis', 'Büro', 'Gebäude', 'Agentur', 'edaktion', 'deutschen', 'Export', 'Antiqu', 'Industrie', 'Apotheke', 'Krankenhaus', 'Werkstatt', 'Stiftung', 'Videothek', 'nstitut', 'Autohaus', 'Arbeitskreis', 'agentur', 'betrieb', 'Betrieb', 'Verwaltung', 'verwaltung', 'Pension', 'direktion', 'Direktion', 'Zentrale', 'zentrale', 'Deutsche', 'Deutscher', 'Deutsches', 'Aktiengesellschaft', 'achhandel', 'oHG', 'OHG'

Wählt man nur wenige Gemeinden aus, so zeigt sich rasch, daß die exportierten Dateien eine Reihe von Nummern enthalten, die in nicht ausgewählten Gemeinden liegen. Hierbei handelt es sich meistens um Querverweise, also z.B. Hauptniederlassungen. Solche Nummern lassen sich mit einem Filterprogramm problemlos beseitigen: Man schreibt nur die Nummern heraus, deren Vorwahl einer der ausgewählten Gemeinden entspricht.

Wählt man hingegen viele oder gar alle Gemeinden aus, so entsteht schon allein durch die Querverweise das Problem von Mehrfacheinträgen der gleichen Telefonnummer. Solche Nummern besäßen eine höhere Auswahlwahrscheinlichkeit. Die einzig saubere Lösung wäre daher eine Löschung der Mehrfacheinträge.

Hierzu benötigt man entweder nach Nummern sortierte Dateien oder eine Häufigkeitsauszählung aller Telefonnummern. Mit Standardsoftware sind beide Aufgaben zumindest für den gesamten Datenbestand mit PCs kaum zu schaffen²². Angesichts dieses Problems wird man eher zu der Annahme neigen wollen, daß die geringe Anzahl von Mehrfacheinträgen und die daraus resultierende Vergrößerung der Auswahlwahrscheinlichkeit dieser Nummern einen vernachlässigbaren Fehler darstellt. Dies gilt um so stärker, da diese Nummern vermutlich eher ausschließlich Geschäftsnummern darstellen, die später ohnehin nicht berücksichtigt werden.

Das Ausmaß der Mehrfacheinträge ist aber erstaunlich hoch, wie etwas mühselige Zählungen zeigen. Im gesamten Datensatz mit ca. 33.606.000 gefilterten Einträgen (ohne Mobil

²² Bei kleineren Datensätzen sind solche Prüfungen natürlich auch mit PCs möglich. Für diese Arbeit wurden die Sortierungen mit einem Mehrprozessor-Unix-System (SGI Power Challenge) durchgeführt; selbst hiermit wurden für 34 Millionen Records mehr als 10 CPU-Minuten benötigt. Die Bestimmung von Doppeleinträgen in der sortierten Datei ist z.B. durch das kurze AWK-Programm DUPS (siehe oben) möglich.

telefone und Datenfehler) sind ca. 32.753.000 Einträge Unikate. 853.000 Einträge finden sich mehrfach (ca. 2.5%). Hierbei handelt es sich um ca. 685.500 verschiedene Einträge. 87% aller Mehrfacheinträge sind Doppeleinträge; 9% kommen dreimal, 2% viermal vor. Die Anzahl der Einträge einer Nummer reicht bis zu 598; wobei die hohen Anzahlen immer Telekom-Service-Nummern entsprechen.

Betrachtet man das Problem der Mehrfacheinträge etwas genauer, so zeigen sich einige Probleme, die zu leicht verzerrten Auswahlen führen können. Nimmt man z.B. alle 928.104 Nummern, die von D-Info 3.0 als "Hamburger" Telefonnummern exportiert werden, so sind ca. 2% der Nummern mehr als einmal vorhanden. Einige Nummern kommen bis zu 30 mal in der Datei vor; allerdings sind 88% der Mehrfacheinträge lediglich Doppeleinträge. Filtert man anhand der oben beschriebenen Kriterien diejenigen Nummern heraus, die vermutlich Geschäftsanschlüsse darstellen, so beträgt der Anteil der mehrfach gelisteten Nummern immer noch 1.1%. Neben verbleibenden Geschäftsanschlüssen sind dies vor allem Paare, die unter zwei verschiedenen Namen eingetragen wurden (93% aller gefilterten Mehrfacheinträge sind Doppeleinträge). Daraus läßt sich der Schluß ziehen, daß z.B. der Anteil der zusammenlebenden, aber unverheirateten Paare in Telefon-Stichproben vermutlich höher geschätzt wird als bei anderen Verfahren.

Stichprobendesignvarianten

Für die Auswahl einer Stichprobe aus einer Telefon-CD gibt es eine Reihe verschiedener Möglichkeiten. Dies gilt in besonderem Maße für Stichproben aus dem gesamten Datenbestand. In diesem Fall liegt es nahe, mehrstufige Auswahlverfahren zu verwenden, da abgesehen von möglichen statistischen Vorteilen mehrstufige Verfahren die Größe der Einzeldateien erheblich reduzieren. Bei allen mehrstufigen Verfahren muß zunächst die Entscheidung getroffen werden, ob als Maß für die Größe einer Primäreinheit (z.B. einer Gemeinde) die Zahl der Einwohner (bzw. der Wahlberechtigten) oder die Zahl der Telefonanschlüsse (bzw. die Zahl der vermutlichen Privatanschlüsse) verwendet werden soll. Beides scheint in der BRD üblich zu sein. Falls die Grundgesamtheit als "allgemeine Bevölkerung" definiert ist, so ist die Verwendung der Zahl der Einwohner bzw. die Zahl der Wahlberechtigten eher korrekt. Wird die Grundgesamtheit allerdings als die in Privathaushalten mit Telefonanschluß lebende Bevölkerung definiert, so ist die Verwendung der Zahl der Anschlüsse korrekt. Wird die Grundgesamtheit wie üblich als "allgemeine Bevölkerung" definiert und trotzdem die Zahl der Anschlüsse als Maß für die Größe der Primäreinheiten verwendet, so muß angenommen werden, daß die Anschlußquote über alle Primäreinheiten konstant ist. Dies ist definitiv falsch. Allerdings wurde bisher weder eine empirische Studie noch eine Simulation zu den tatsächlichen Konsequenzen dieser Vorgehensweise publiziert. Hier könnte eine Ursache für mögliche Unterschiede in den Ergebnissen verschiedener Datenerhebungsinstitute liegen. Am einfachsten ist natürlich die Verwendung der Zahl der An-

schlüsse als Maß für die Größe der Primäreinheiten, da dann alle notwendigen Daten prinzipiell der CD entnommen werden können. Im folgenden werden daher - ungeachtet der statistischen und inhaltlichen Konsequenzen - zunächst Verfahren auf der Basis dieser Vorgehensweise erörtert.

Geschichtete Stichproben

Besonders naheliegend ist die Verwendung einer nach Bundesländern geschichteten Stichprobe. Die D-Info 3.0 CD erlaubt zwar indirekt den Export nach einzelnen Bundesländern²³, die resultierenden Dateien sind allerdings fehlerhaft. Vergleicht man die Zahl der Anschlüsse der einzelnen Bundesländerdateien mit der Gesamtzahl der Anschlüsse, so ergibt die zusammengeführte Bundesländerdatei 37.1 Millionen Anschlüsse, die Gesamtdatei enthält hingegen nur 34.3 Millionen (ungefilterte) Anschlüsse. Folglich werden 2.8 Millionen Anschlüsse mehrfach exportiert. Die Ursache liegt vor allem in der Zuordnung Hamburgs zu Schleswig-Holstein und zu Niedersachsen sowie Bremens zu Niedersachsen. Korrigiert man diese gravierenden Fehlzuordnungen, so verbleiben immer noch 553.000 überzählige Einträge. Zu ähnlichen Zahlen gelangt man über die Analyse der Vorwahlnummern. Bereinigt man die exportierten Dateien der einzelnen Länder um die Sondernummern (0130, 0161, 0190 etc.) so zeigen sich 5247 verschiedene Vorwahlnummern, davon 5198 mit mehr als 10 Anschlüssen. 67 Vorwahlen mit mehr als 1000 Anschlüssen erscheinen mehrfach. Insgesamt finden sich fast 561.000 Einträge in mehr als einem Bundesland. Dies können Querverweise sein (wie z.B. auf Hauptniederlassungen), aber auch Vorwahlen in mehr als einem Bundesland. Beispiele für "berechtigte" Mehrfachzuordnungen sind einerseits Mannheim und Ludwigshafen, andererseits Ulm. Es ergeben sich aber auch schlicht falsche Mehrfachzuordnungen, z.B. bei 02224 (Bad Honnef) und 0241 (Düren). Zusammenfassend kann daher festgestellt werden, daß sich zwar aus den Daten der D-Info 3.0 CD eine nach Ländern geschichtete Stichprobe gewinnen ließe, allerdings nur mit sehr hohem Arbeitsaufwand. Für praktische Stichproben empfiehlt sich daher ein anderes Verfahren.

PPS-Stichproben

Eine andere Möglichkeit der Auswahl für eine bundesweite Studie besteht darin, die CD in gleicher Weise wie normale Telefonbücher als Auswahlgrundlage zu verwenden. In diesem Fall würde man vermutlich zu einem PPS-Design ("probability proportional to size", vgl. *Schnell, Hill, Esser* 1995:268-270) greifen: Auswahl der Primäreinheiten (z.B. Gemeinden oder Vorwahlen) entsprechend ihrer Größe, innerhalb der Primäreinheiten erfolgt dann die

23 Der Export einzelner Bundesländer ist nicht direkt möglich, sondern dadurch, daß man zunächst willkürlich einen Ort des jeweiligen Bundeslandes auswählt und dann im folgenden Exportdialog des Abfrageprogramms der D-Info CD die entsprechende Option selektiert.

Auswahl einer konstanten Zahl von Sekundäreinheiten (Haushalte) über eine einfache Zufallsstichprobe. Praktisch lassen sich PPS-Stichproben entweder mittels kommerzieller Software (*Frankel, Spencer* 1990), Makros in Datenanalysesystemen (*Lehtonen, Pakhinen* 1994:55-56) oder auch durch einige Zeilen AWK mittels des sogenannten Kumulationsverfahrens (vgl. *Schnell, Hill, Esser* 1995:445-447) ziehen²⁴. Das Problem bei PPS-Stichproben liegt darin, daß man die Größe der Primäreinheiten benötigt. Die Zahl der zu einer Vorwahlnummer gehörenden Einheiten läßt sich aus der D-Info 3.0 CD auszählen, solange es sich um kleinere Datenbestände handelt. Wie oben geschildert, gibt es hier zwar das Problem der Mehrfacheinträge, daß man aber wohl bei den meisten Fragestellungen ohne große Folgen ignorieren kann. Zumindest beim gesamten Datensatz wirft aber die Häufigkeitsauszählung des gesamten Datenbestandes größere technische Probleme auf²⁵. Falls man nicht die Zahl der Anschlüsse als Maß für die Größe der Primäreinheiten verwenden will, so kann man z.B. auf die Einwohnerzahlen der Gemeinden oder wie beim ADM-Design auf die Größe der Wahlbezirke rekurren²⁶. Da das ADM-Ziehungsband nicht öffentlich zugänglich ist, müßten die Unterlagen über die Größe der Stimmbezirke beschafft werden. Eine Datei mit diesen Angaben für alle 80.000 Wahlbezirke ist beim Bundeswahlleiter im Statistischen Bundesamt gegen eine Gebühr von ca. 4.000 DM erhältlich. In einem weiteren Arbeitsschritt müßten kleinere Wahlbezirke zusammengefaßt und dann mit den Vorwahlen für die zusammengefaßten Einheiten zusammengeführt werden. Beides wäre mit den Aggregierungs- und Filemergeroutinen von Datenanalysesystemen wie SPSS oder STATA mit jeweils einem Kommando prinzipiell möglich²⁷. Ähnliches gilt für die Einwohnerzahlen der Gemeinden, wobei nur für die größeren Gemeinden diese Zahlen problemlos beschaffbar sind.

Nachdem im ersten Schritt diese Vorwahlnummern mittels PPS ausgewählt wurden, muß im zweiten Schritt eine fixe Zahl von Einträgen für jede Vorwahl ausgewählt werden. Es empfiehlt sich, die Zahl der Einträge pro Vorwahl klein zu wählen, um den Klumpen-Effekt gering zu halten. Eine **realisierte** Zahl von weniger als 10 Einträgen bei ca. 250 verschiedenen Vorwahlen wäre eine übliche Größe.

24 Die letzte Methode hat den Vorteil, daß man mit nichtaggregierten Datenfiles beliebiger Größe arbeiten kann. Außerdem ist das Programm kostenneutral. Die PPS-AWK-Implementierung in DOS ist allerdings etwas mühsam, da AWK keine Sortieroutine enthält sondern auf das Betriebssystem zurückgreift.

25 Mit den Standardprogrammen ist diese Aufgabe nicht zu leisten. Mit Turbo-Pascal läßt sich die Gesamtdatei nicht bearbeiten. In AWK könnte man dieses Problem mit assoziativen Arrays bearbeiten; dies scheitert aber an einem Memory-Limit dieser Arrays in MSDOS-AWK. Am einfachsten läßt sich dieses Problem mit einem C oder FORTRAN-Compiler angehen.

26 Diese Anlehnung an das ADM-Design erwähnt auch *Zeh* (1987:344).

27 Da die Stimmbezirksdatei die Gemeindekennziffer enthält, müßten alle Vorwahlnummern für alle einzelnen Gemeindekennziffern in einer Datei vorliegen. Eine solche Datei scheint nicht allgemein verfügbar zu sein.

Liegt eine Datei vor, in der nach Vorwahlen sortierte Telefonnummern vorhanden sind, lassen sich die Einträge z.B. entweder durch eine einfache Zufallsstichprobe oder durch ein systematisches Auswahlverfahren (z.B. Ziehen jedes n-ten Eintrags) ziehen. Beide Methoden erfordern nur wenige AWK-Zeilen (siehe weiter unten). Liegt keine solche Datei vor, bleibt als praktisch gangbarer Weg zunächst nur der klassische Weg über die gedruckten Telefonbücher. Dann muß für jede PPS-ermittelte Vorwahl wieder manuell mit einem systematischen Verfahren aus dem Telefonbuch gezogen werden. Hierbei muß beachtet werden, daß das Druckbild der Telefonbücher sich unterscheidet, so daß die Anzahl der Einträge pro Seite unterschiedlich ist. Korrekt wäre ein systematisches Verfahren also erst dann, wenn die Zahl der Einträge pro Vorwahl insgesamt bekannt oder jeweils geschätzt würde²⁸. Interessanterweise gibt aber die Telekom-CD die Zahl der Einträge pro Vorwahl in der Bildschirmmaske nach Auswahl eines Ortes aus. Damit kann diese CD direkt als Basis eines manuell durchgeführten systematischen Auswahlverfahrens dienen; allerdings muß man für dieses Verfahren den Schreibtisch nicht mehr verlassen.

Übersicht 3: AWK-Programm SAMPLE.AWK: Zieht eine Zufallsstichprobe der Größe n aus N Eingabezeilen

```
# SAMPLE.AWK: select a random sample
BEGIN {
  if (ARGC != 4) {
    print "Syntax:  AWK sample filename N n "
    print "where N=number of input  lines"
    print "      n=number of output lines"
    print "Example: AWK sample frame.lst 2000 10"
    exit }
  ni=0; srand(); n=ARGV[2]; m=ARGV[3]; ARGC=2 }
{ if ( rand() < (m-ni)/(n-NR+1) ) { print $0 ; ni=ni+1 } }
END { if ((NR != n) && (NR > 1)) \
  print "Number of units is ",NR," not ",n,\
        ". Selection invalid." }
```

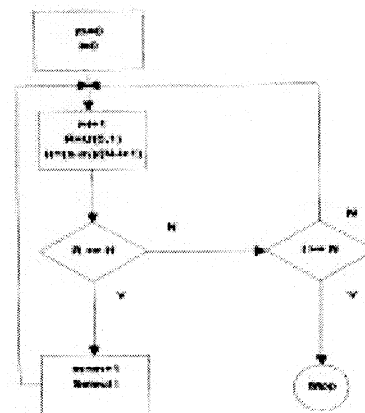
Sollten zukünftig die CDs diese Bestimmung der Zahl der Einträge pro Vorwahl nicht mehr ermöglichen, wäre die Ziehung weiter erschwert und mit erheblichen Aufwand verbunden²⁹. Aufgrund der Kosten und des Arbeitsaufwandes scheiden alle PPS-Verfahren, die nicht auf der Auszählung der CD basieren, daher zumindest für "Lehrforschung" praktisch aus. Verwendet man aber die CD-Zählung als Basis der PPS-Stichprobe, so muß man (wie oben erwähnt) entweder die Grundgesamtheit ungewöhnlich definieren oder die unklaren Konsequenzen der falschen Annahme konstanter Anschlußquoten tragen. Aus diesem Grund

²⁸ vgl. *Frey, Kunz, Lüschen* (1990:87).

²⁹ Erfahrungsgemäß brauchen zwei studentische Hilfskräfte ca. 1 Woche, um auf diese Art eine Zufallsstichprobe von 2000 Zahlen zu ziehen. *Frey, Kunz, Lüschen* (1990:87) erwähnen ohne Seitenangabe den Vorschlag von *Blankenship* (1977), die Auswahl durch eine Schablone vorzunehmen. Bereitet man entsprechend für jedes der vorliegenden Telefonbücher eine eigene Schablone vor, so ist der Auswahlvorgang wesentlich rascher durchführbar.

scheint die einstufige Auswahl aus dem gesamten Datenbestand für bundesweite Surveys die einfachste, schnellste und am wenigsten problematische Lösung darzustellen. Für regional begrenzte Studien ist dies ohnehin fast immer die Methode der Wahl.

Abbildung 2: Sequentielle Ziehung ohne Zurücklegen (*Kennedy, Gentle* 1980:238)



Einstufige Auswahlverfahren

Ein einfacher Algorithmus zur Auswahl von exakt n Elementen aus einer Liste von N Elementen findet sich bei *Kennedy, Gentle* (1980:238, vgl. Abbildung 1)³⁰. Das entsprechende AWK-Programm SAMPLE.AWK findet sich oben³¹. Diese Methode setzt voraus, daß die Anzahl (N) der Elemente der Grundgesamtheit vor der Ziehung bekannt ist. Meist läßt sich diese Zahl recht einfach bestimmen (z.B. mit dem AWK-Oneliner "LINES.AWK", siehe Text 1). Bei großen Datenmengen wie beim gesamten Datenbestand der D-Info CD kann die Ermittlung von N aber mit Schwierigkeiten behaftet sein. In solchen Fällen empfiehlt sich die Verwendung eines Algorithmus, der nicht die Kenntnis von N voraussetzt³². Aufgrund der unterschiedlichen Ausfallursachen bei Telefonsurveys (vgl. *Schnell* 1997:116-

30 Im Flußdiagramm ist i der Laufindex, m ist die Zahl der bisher ausgewählten Einheiten, n die Gesamtzahl der Einheiten in der Auswahlgrundlage. $U(0,1)$ bezeichnet einen Pseudo-Zufallszahlengenerator, der gleich verteilte Zahlen im Intervall zwischen Null und Eins produziert. Dieser Algorithmus wird auch in einer Reihe kommerzieller Statistikpakete verwendet.

31 Da die in den üblichen Compilern eingebauten Pseudo-Zufallszahlengeneratoren meist kaum dokumentiert sind (vgl. hingegen *Dohmann, Falk, Lessenich* 1991) und auch die besten Generatoren mit Problemen behaftet zu sein scheinen (vgl. *Ferrenberg, Landau, Wong* 1992), empfiehlt sich im allgemeinen eher die Verwendung explizit programmierter Generatoren. In vielen Anwendungen hat sich bisher z.B. der portable Pseudo-Zufallszahlengenerator von *Park, Miller* (1988) bewährt. Das Programm findet sich z.B. bei *Schnell, Hill, Esser* (1995:443). Programmiert man den Generator unter AWK, so wird das Programm etwas unübersichtlich (Verwendung von Subroutinen) und langsam. Für praktische Zwecke würde man dieses Programm daher eher in einer allgemeinen Programmiersprache schreiben.

32 Ein Beispiel hierfür findet sich in der Arbeit von *Pinkham* (1987).

127) sollte eine wesentlich höhere Anzahl von Nummern (n) ausgewählt werden, als realisiert werden soll. Nach den Erfahrungen in einer Reihe von RLD-Surveys der "allgemeinen Bevölkerung" sollte diese Zahl bei ca. 300% der angestrebten Stichprobengröße liegen.

Die Zukunft der Telefonsurveys auf der Grundlage der Telefon-CDs

Das Erscheinen der Telefon-CD D-Info 3.0 hat die Möglichkeiten der Stichprobenziehung für Telefonsurveys vereinfacht. Durch den Export einer ASCII-Datei für das Untersuchungsgebiet läßt sich mit den hier angegebenen Programmen und Hilfsmitteln relativ problemlos eine Zufallsstichprobe der eingetragenen Telefonnummern ziehen bzw. eine RLD-Stichprobe erzeugen. Liegt einmal eine bereinigte ASCII-Datei der Telefonnummern vor, so kann eine Stichprobe in wenigen Minuten gezogen werden. Leider ist aufgrund von Datenschutzbemühungen zu erwarten, daß eine solche Auswahlgrundlage nicht dauernd zur Verfügung stehen wird. In diesem Fall bleiben nur wenige Möglichkeiten:

- Entweder man rekurriert zumindest für die Generierung der Ausgangsnummern des RLD-Verfahrens wieder auf die gedruckten Telefonbücher;
- man verwendet die alten CDs als Ausgangspunkt des RLD-Verfahrens
- oder man verwendet die Länge der eigentlichen Telefonnummer (ohne Vorwahl) als Ausgangspunkt eines "Random-Digit-Dialings" (RDD), bei dem alle Ziffern aus einem Pseudo-Zufallszahlengenerator gezogen werden.

Der Rückgriff auf die gedruckten Telefonbücher ist mühsam, aber sicher auch in Zukunft möglich³³. Die weitere Verwendung der alten CDs müßte voraussetzen, daß sich die Häufigkeiten der Anfangsdigits des Telefonnetzes nicht verändert. Die Verwendung der Länge der Telefonnummern (ohne Vorwahl) als Ausgangsbasis für eine vollständige Zufallsgenerierung der Nummern setzt die Konstanz der relativen Häufigkeit unterschiedlicher Telefonnummernlängen voraus. Diese wird sich vermutlich langsamer ändern als die relative Häufigkeit der Anfangsdigits. Damit wäre RDD gegenüber RLD auf der Basis älterer CDs geraume Zeit verwendbar - bis entweder neue CDs verfügbar werden oder die Telekom (endlich) Daten über die Häufigkeit einzelner Anfangsdigits und die Länge der Telefonnummern zur Verfügung stellt. Allerdings wäre es auch möglich, daß ein immer größerer Anteil der Telefonbesitzer dazu übergeht, ihre Anrufbeantworter zum Aussieben der Anrufer zu verwenden - falls sie nicht ohnehin ihre intelligenten ISDN-Telefone so programmieren, daß sie keine Anrufe von unbekannten Anrufern akzeptieren. Dies wäre das definitive Ende der Telefonsurveys als Zugangsmethode zur "allgemeinen Bevölkerung".

33 Solange die Telekom-CD die Anzahl der Einträge pro Vorwahl ausgibt, kann diese CD zusammen mit der Stimmbezirksdatei als Grundlage für ein auf der letzten Stufe manuell durchgeführtes PPS-Verfahren verwendet werden (siehe oben).

Literatur

- Aho, A.V., Kernighan, B.W., Weinberger, P.J.** (1988):
The AWK Programming Language, Reading/Mass. (Addison-Wesley).
- Blankenship, A.B.** (1977):
Professional Telephone Surveys, New York (McGraw-Hill).
- Borland GmbH** (1992):
Turbo Pascal 7.0, München.
- Chapman, G.** u.a. (1990):
Monty Pythons Flying Circus: Just the Words, Vol.2, London.
- Dohmann, B., Falk, M., Lessenich, K.** (1991):
The Random Number Generators of the Turbo Pascal Family; in: Statistical Software Newsletter, 12, 1, S.129-132.
- Ferrenberg, A.M., Landau, D.P., Wong, Y.J.** (1992):
Monte Carlo Simulations: Hidden errors from "good" random number generators; in: Physical Review Letters, 69, 23, 3382-3384.
- Frankel, M.R., Spencer, B.D.** (1990):
Sample: A Supplementary Module for Systat and Sygraph, Evanston, Ill. (SYSTAT, Inc.).
- Frey, J.H., Kunz, G., Lüschen, G.** (1990):
Telefonumfragen in der Sozialforschung, Opladen.
- Fuchs, M.** (1994):
Umfrageforschung mit Telefon und Computer, Weinheim (Beltz).
- Häder, S.** (1996):
Wer sind die Nonpubs? Zum Problem anonymer Anschlüsse bei Telefonumfragen; in: ZUMA-Nachrichten 39, S.45-68.
- Hutchinson, H.** (ed.) (1996):
Ci-3 User Manual, Version 1.1, SAWTOOTH Software, Sequim, WA.
- Kennedy, W.J., Gentle, J.E.** (1980):
Statistical Computing, New York (Marcel Dekker).
- Lavrakas, P.J.** (1987):
Telephone Survey Methods, Beverly Hills (Sage).
- Lehtonen, R., Pahkinen, E.J.** (1994):
Practical Methods for Design and Analysis of Complex Surveys, Chichester (Wiley).
- Marhenke, W.** (1996):
Telefonanschlußdaten als Auswahlgrundlage, Vortrag anläßlich des ZUMA-Symposiums "Vergleich von Stichprobenverfahren", Mannheim; wird 1997 in einer ZUMA-Buchpublikation erscheinen.
- Park, S.K., Miller, K.W.** (1988):
Random Number Generators: Good Ones are Hard to Find; in: Communications of the ACM, October 1988 (Volume 31, Number 10).
- Pinkham, R.S.** (1987):
An Efficient Algorithm for Drawing a Simple Random Sample; in: Applied Statistics, 36, 3, S.370-372.
- Schnell, R.** (1997):
Nonresponse in Bevölkerungsumfragen, Opladen (Leske+Budrich).
- Schnell, R., Hill, P.B., Esser, E.** (1995):
Methoden der empirischen Sozialforschung, 5. Auflage, München (Oldenbourg).
- Wall, L., Schwartz, R.L.** (1993):
Programmieren in perl, München (Hanser).
- Zeh, J.** (1987):
Stichprobenbildung bei Telefonumfragen; in: Angewandte Sozialforschung, 14, 4, S.337-347.